Deep Learning Internship Report

Dr.Sanjay Kumar Singh* and Ajeya Krishna†

Indian Institute of Technology (BHU), Varanasi, U.P. India - 221005

I. ABSTRACT

The COVID-19 viral outbreaks have exceeded our expectations and have shattered all prior records for virus outbreaks. Corona virus infection creates a dangerous sickness that can lead to death due to significant alveolar damage and progressive respiratory collapse. The application of computer vision technology to detect and classify this virus from a chest X-ray picture can be a very valuable addition to the less sensitive traditional method of detecting COVID-19, namely reverse transcription polymerase chain reaction (RT-PCR).This automated technique has the potential to improve on current COVID-19 treatment methods while also alleviating the scarcity of skilled physicians in rural areas. Again, segmenting diseased regions from a chest X-ray picture can aid medical practitioners in gaining insight into the afflicted area. So, in this research, we employed a deep learning-based transfer learning approach for CT-scan and X-ray image classification, and a U-Net architecture for segmentation to segment the afflicted region. On the available X-ray dataset, 99.7% classification accuracy was obtained, 99.4% on the available CT-scan dataset, and 87% average accuracy from the segmentation process.

II. INTRODUCTION

In December 2019, the new corona virus (COVID-19) spreads from Wuhan, China, to the rest of the world.First, it assaults the lungs and respiratory system of a human body, earning it the designation (SARS-COVID-19), which stands for severe acute respiratory syndrome. As a result, screening mechanisms for this organ, such as X-rays and Computer Tomography (CT) images, can play an essential role in successfully recognizing disorders and taking appropriate steps. With the use of Computer Vision and Deep Learning technologies, this approach may also be used to facilitate the creation of curative medicine. Again, segmenting the affected region from this organ can aid medical workers in providing early treatment, care, and isolation in order to prevent the infection from spreading.

*sks.cse@iitbhu.ac.in

†ajeya.krishnap.phy17@iitbhu.ac.in

Medical radiography imaging is used as a supplement to the RT-PCR test to prevent the devastating infection of COVID-19. This is predicated on the fact that most COVID-19 patients had clinical symptoms of lung infection on chest X-rays . In addition, a CT scan is required to monitor the severity of the sickness. CT scan tests, on the other hand, have a serious constraint in terms of diagnostic time: even expert radiologists take roughly 21.5 minutes to review each case's test data. During the pandemic epidemic, skilled radiologists are in short supply, making it difficult to detect possibly infected patients in a timely manner. As a result, automated diagnosis systems are in great demand.

III. DATASETS

We have utilized different datasets for classification and segmentation. For classification, we have used both Covid and Non-Covid pictures of X-ray and CT-scans. We used 13816 images for X-ray, consisting of 3616 Covid-19 and 10200 Non-Covid images. We also used 14500 images for CT-scans, consisting of 7593 Covid-19 and 6893 Non-Covid images.For segmentation, we have used 2 datasets.Both the datasets consists of CT-scans along with their infection masks. In the first dataset, it consists of 1200 images with the infection mask. In the second dataset, it consists of 1564 CT-scans along with their infection masks. We considered 70% of the datasets used for training, 10% for validation and 20% for testing purpose. To extract the most out of the data in this study, we applied a variety of data preparation approaches. We have applied variant of Adaptive Histogram Equalization (AHE) called as CLAHE [1] to improve the contrast of the images. We reduced images to the core section to concentrate on the primary component.

Dataset	Covid-19	Normal
CT-scans [2]	7593	6893
X-ray [3]	3616	10200

TABLE I: CT-scans and X-ray Datasets

Dataset	Mask Images
Dataset-1 [4]	1200
Dataset-2 [5]	1564

TABLE II: Infection Mask Datasets

IV. METHODOLOGY

The overall methodology is given in Fig. 1. For classification and segmentation, we used two different approaches. We employed an uncertainity aware transfer learning approach for classification. We utilized the U-net framework for segmentation. Starting with classification, the first step is processing the data. We have removed the images which are blurry, not clear and not useful. Then, we applied Adaptive Histogram Equalization (AHE) called as CLAHE [1] to improve the contrast of the images and cropped the images to the core part.

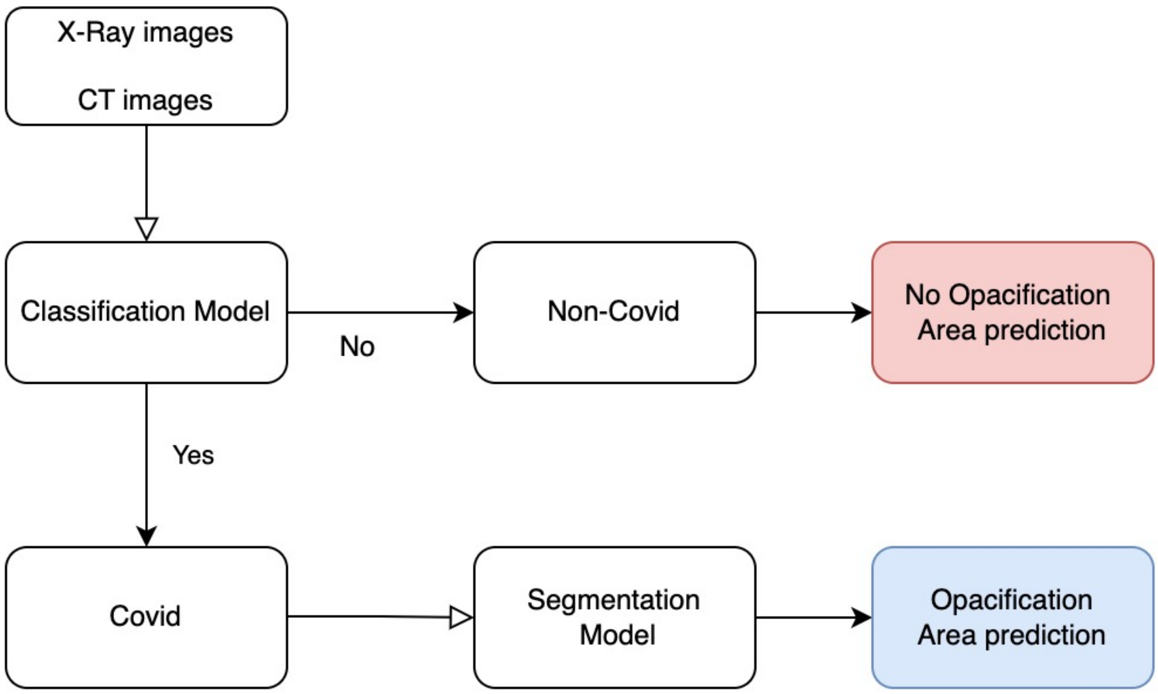


FIG. 1: The flow chart of our suggested methodology.

A. Classification

The transfer learning technique will be used to train machine learning models for COVID-19 identification. Two significant factors drive our decision to use a transfer learning architecture to tackle the COVID-19 detection problem:

- Training DNN/CNN models necessitates a large quantity of data, which is not feasible for COVID-19 because the number of captured and tagged photos is often in the hundreds.
- DNN/CNN model training is computationally intensive. Even if there are dozens or millions of photos available, it's still a good idea to test the utility of existing pretrained models for data representation and feature extraction first.

To detect the presence of COVID-19, the suggested framework only relies on the information content of X-rays and CT images. In this report, we evaluate two best pretrained networks on the ImageNet data set and import and adjust them for COVID-19 identification. ResNet50 and InceptionResNetV2 are the names of these networks. All of these networks have demonstrated cutting-edge performance in accurately identifying photos from the ImageNet data set. Because these networks have numerous layers and millions of trainable parameters, training them is computationally intensive. The proposed framework's key claim is that there are basic parallels between image detection/recognition tasks and the COVID-19 binary classification issue utilizing pictures. As a result, learnings from the former may be securely transferred to the latter to reduce the training period. While all two pretrained networks were created using non-medical pictures, it is plausible to believe that their alteration of X-ray and CT image pixels would aid in categorization.

As shown in Fig. 2, during the training phase, the convolutional layer parameters are maintained locked. For hierarchical feature extraction, the convolutional layers of these two pretrained models are fed by X-ray and CT images. Different machine learning classifiers replace the front end of the pretrained networks to differentiate Covid and non-Covid situations. It's worth noting that the pooling process is skipped in the pretrained networks' final convolutional layer. This is to prevent informative characteristics from being lost before they are passed on to categorization models.

We also studied the uncertainity associated with the classification models.There are two types of uncertainties, which needs to be considered for deep learning models.

1. Aleatoric uncertainty is a type of uncertainty that is caused by noise in the data generation process. This kind of unpredictability can't be eliminated.

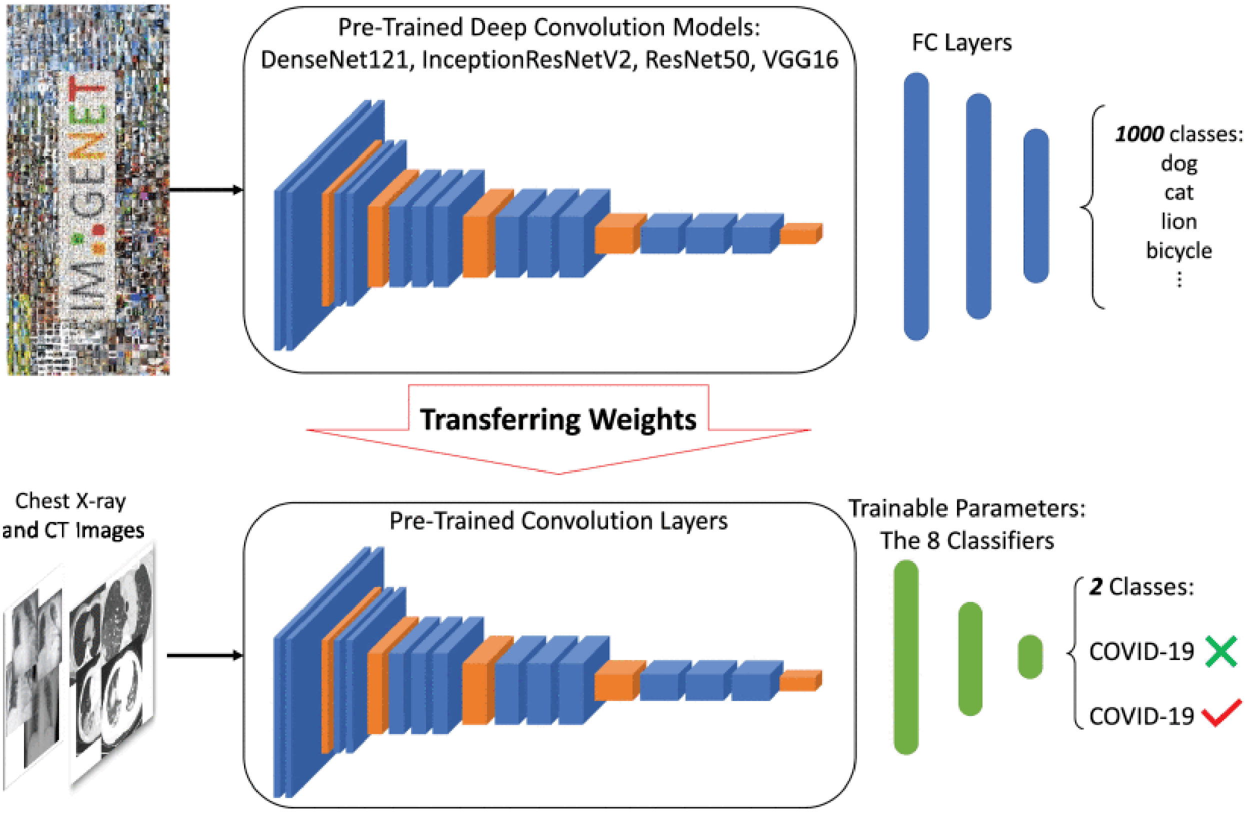


FIG. 2: The suggested transfer learning-based architecture for COVID-19 detection utilizing X-ray and CT images is depicted in a block diagram.

2. The term epistemic uncertainty refers to the lack of knowledge about the model. Unlike aleatoric uncertainty, epistemic uncertainty may be reduced by collecting more training samples from a variety of contexts.

We will concentrate on epistemic uncertainty in this article since it is intimately related to the generalization capacity of models for fresh data. The epistemic uncertainty might be calculated using the prediction variance. As a measure of epistemic uncertainty, we calculate prediction entropy.

As a measure of epistemic uncertainty, we calculate prediction entropy. The prediction entropy is a tool for assessing the degree of uncertainty in scores provided by various models. The entropy of the mean predictive distribution is used to compute the ensemble epistemic uncertainty (by averaging all predicted distributions)

$$\hat{p}(y|x) = \frac{1}{N}\sum_{i=1}^{N} p_{\theta i}(y|x) \tag{1}$$

$$H(\hat{p}(y|X)) = \sum_{i=0}^{C} \hat{p}(y_i|x) log \hat{p}(y_i|x) \tag{2}$$

where θ_i is the set of parameters for the i th network element and C ranges over all classes. For instance, suppose that for a given input, an individual neural network predicts that the input belongs to class 1 with x amount of probability and to class 0 with y amount. If we repeat this procedure ten times for that specified input, it is similar to ensembling ten networks for predicting the output probability. The final output probability can be calculated using (1). Now, imagine the average probability predicts that an input belongs to class 1 and 0 with 0.6 and 0.4, respectively. Based on (2) the prediction entropy can be calculated as $0.6 * log(0.6) + 0.4 * log(0.4)$. It is obvious that the prediction entropy becomes zero when the output is assigned to a class with high probability and becomes maximum when the network is uncertain about its outcome.

B. Segmentation

The general approach to semantically segmenting pictures is to develop a structure that collects features through consecutive convolutions and outputs a segmentation map.

U-NET was created with the intention of comprehending and segmenting medical images. It is a significant architecture in the medical imaging automation society and has a wide range of applications in the sector. In this part, we go through the network's core technological aspects and how they contribute to successful outcomes.

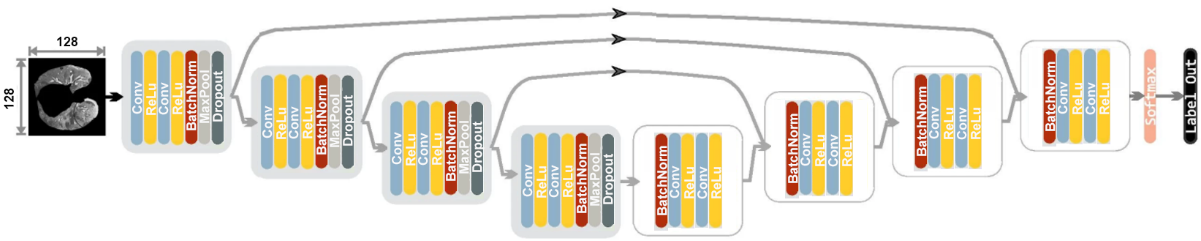


FIG. 3: Flow Diagram of the U-Net model that we have used

This network's design is divided into three parts: encoding, decoding, and attention block. Following ReLU layers, the encoding route consists of numerous patches of convolutions with filters of size 33 and unity strides in both directions. The encoder network consists of 4 blocks. In each block we would downscale the image. Each block consists of Convolutional layer, ReLU activation, Max Pooling layer, Batch Normalization, Dropout. This route extracts the input's essential features and returns a feature vector of a certain length. The second route network is known as decoding network. This would scale up the image that was downscaled by the encoding network. This decoding network consists of 4 block. In each block, it has Batch Normalization, Convolutional layer, ReLU activation. The second route extracts data from the encoding path by copying and cropping, as well as from the feature vector using up-convolutions, and forms an output segmentation map via a series of operations. The operation that connects the first and second pathways is a crucial part of this system. This connection enables the network to obtain very precise information from the encoding path, resulting in a segmentation mask that is as near to the intended output as feasible.

We have added attention block in this U-Net model. The attention block consists of Convolution layer for both decoder and encoder, activation layers. These attention blocks are placed in each skip connections that connects encoder and decoder.

Talking about the loss function, we have implemented weighted boundary loss and binary cross entropy dice loss. Using only binary cross entropy or dice loss for these highly imbalanced segmentations would not yield good results. Such localised summations have values that differ by many orders of magnitude across classes for extremely imbalanced segmentations, affecting training performance and stability.

Apart from the U-Net model training, we have also used Transfer Learning technique in the segmentation. We have considered EfficientNetB[0-7]. We follow the same technique in this section as we did in the previous section. The last layer was removed, and the picture's features were retrieved. We then used standard neural networks to train the model to adapt to the features.

V. EXPERIMENTS

In this section we present our different experiment setups that we have considered. To start with we present the experiment considerations in the classification model. We give brief introduction of the pre-trained models.

1. **ResNet50**

   Residual convolutional network (ResNet) is one of the most popular deep structure, which is used for classification problem (winner of ImageNet competition in 2015). Residual blocks enable the network to provide a direct path to its early layers. This helps the gradient flow easily in the backpropagation algorithm.

2. **InceptionResNetV2**

Szegedy et al. presented a novel structure that helps to go deeper through convolution networks. Deep networks are prone to overfitting. They solve this solution using inception blocks. Furthermore, they use residual blocks and create InceptionResNetV2, which uses the combination of residual and inception blocks wisely.

Architecture	Paper	Input Size	Number of Features	Number of Parameters
ResNet50	2015	224 X 224	1,00,352	2,35,87,712
InceptionResNetV2	2015	299 X 299	98,304	5,43,36,736

TABLE III: Information About Two Considered Architectures for Transfer Learning.

High-level information about these pretrained models is shown in Table III. As shown in Fig. 2, network weights are kept frozen during the transfer learning procedure. The size of our input images is 224×224 for ResNet50. The input size for the InceptionResNetV2 architecture is 299×299 .

The COVID-19 detection is a binary classification problem where the input is an image (chest X-ray or CT image) and the output is a binary label representing the presence or absence of COVID-19. Here, images are first processed by the convolutional layers of two pretrained networks. Hierarchically extracted features are then processed by multiple classifiers. We use eight classifiers to process features: k-nearest neighbors (kNNs), Polynomial support vector machine (linear SVM), Radial basis function (RBF) SVM, Random Forest (RF), Multilayer Perceptron (NN), Adaboost, Naive Bayes, and XG Boost.

Coming to Segmentation model, we state of cosnideration here. We took EfficientNetB[0-7] for transfer learning technique and general deep U-Net for another dataset.

- **EfficientNet**

  EfficientNet is a convolutional neural network design and scaling approach that uses a compound coefficient to scale all depth/width/resolution dimensions evenly. The EfficientNet scaling approach consistently increases network breadth, depth, and resolution with a set of preset scaling coefficients, unlike standard practise, which adjusts these parameters randomly.

We followed the same technique as mentioned in the Fig. 2. Coming to the U-Net part, we considered various hyper-parameters. The hyper-parameters includes Block size, Epochs, Batch Size, Learning Rate.

Block Size	Epochs	Batch Size	Learning Rate
4	100	32	1*10ˆ-5
6	200	64	5*10ˆ-5
	300	128	Cosine Annealing Learning Rate Scheduler
	400		
	500		

TABLE IV: Table representing different hyper-parameters. We tried all the combination of these 4 different hyper-parameters.

VI. EVALUATION METRICS

For the Evaluation of the models, we have used various metrics that conveys the performance of the model. The metrics are:

1. Accuracy
2. F1-score
3. Recall
4. Precision
5. Specificity
6. Dice Coefficient
7. Jaccard Index

A confusion matrix is a table that is frequently used to describe the performance of a classification model on a set of test data with known true values. It contains TP (True Positive),TN (True Negative),FP (False Positive),FN (False Negative).

Accuracy : The simplest intuitive performance metric is accuracy, which is just the ratio of properly predicted observations to all observations.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

Precision : Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall/Sensitivity : Recall is the ratio of correctly predicted positive observations to the all observations in actual class.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Specificity : It is also known as True Negative Rate. It refers to the proportion of those who received a negative result on this test out of those who do not actually have the condition.

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

F1-score/Dice coeff : The weighted average of Precision and Recall is the F1-score. As a result, this score considers both false positives and false negatives.

$$F1score = \frac{2TP}{2TP + FN + FP} \tag{7}$$

Jaccard Index : It is also known as Intersection over Union. It is the ration of True positive to sum of True positive, False Negative and False Positive.

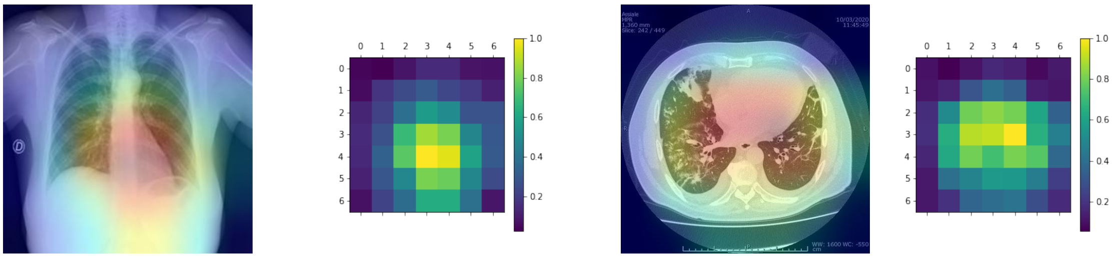
$$IoU = \frac{TP}{TP + FN + FP} \tag{8}$$

VII. SIMULATIONS AND RESULTS

In this section we present our simulations and results of both Classification and Segmentation. We tested our Classification Model on 2 different datasets. The results are attached in the Classification section. We tested our Segmentation Model on 2 different datasets using 2 different approaches.The results are attached in the Segmentation section.

We use the Grad-CAM idea to demonstrate the efficacy of extracted features. Because it requires certain inputs and produces some outputs in such a way that no one knows how it works, neural network design is sometimes referred to as a "black box." Gradient-weighted class activation mapping (Grad-CAM) provides visual explanations of how our model makes decisions (spatial information obscured by layers). This may be calculated by backpropagating the target gradients via the convolutional layers, resulting in a heatmap. The most significant locations of input for the categorization choice will be highlighted in this heatmap.

We use the same structure that can locate the most relevant pixels in a typical image to forecast a given label to obtain insight into our suggested model. Figure 4 depicts the Grad-CAMs and heatmaps of four photos, demonstrating how our ResNet50 functions.



(a) (b) (c) (d)

FIG. 4: Grad-CAMs and Heatmaps show how our model makes decision. (a) Grad-CAM of X-ray. (b) Heatmap of X-ray. (c) Grad-CAM of CT-scan. (d) Heatmap of CT-scan.

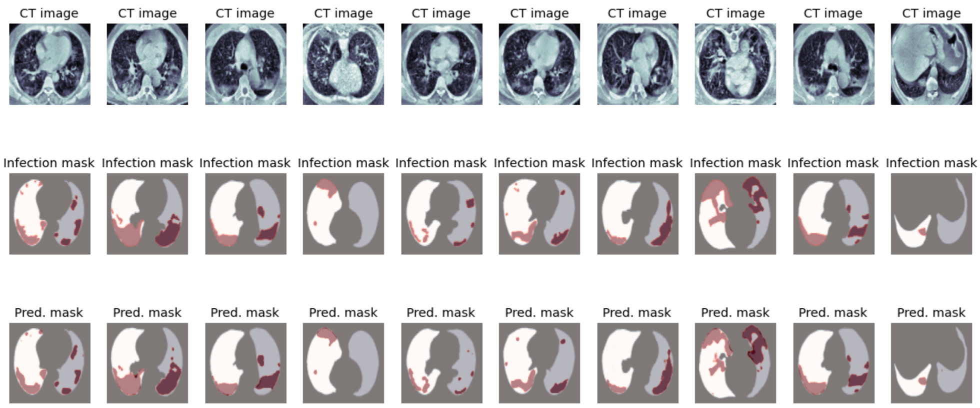


FIG. 5: Comparison of the test results with the original snapshots.

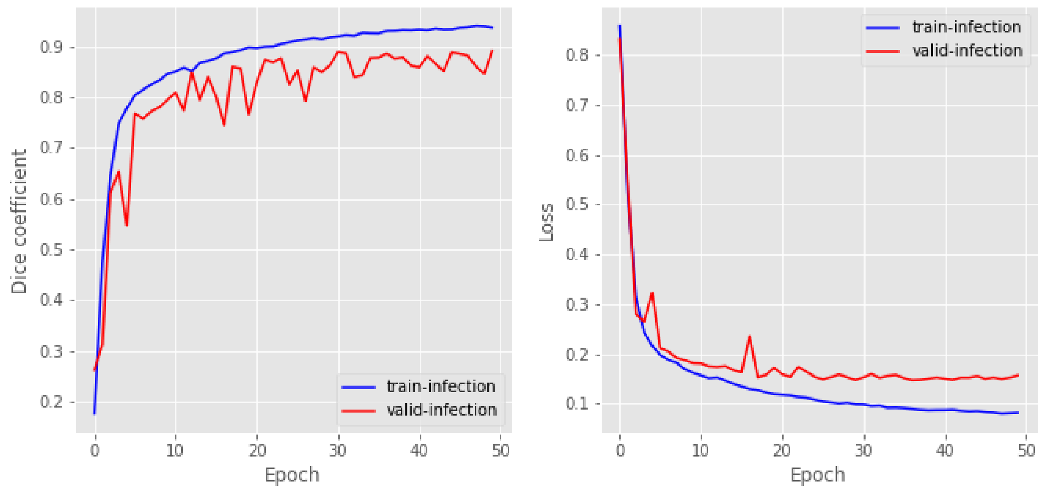


FIG. 6: Dice Coefficient vs Epoch and Loss vs Epoch of U-Net Attention model which iterated for 50 epochs, 32 batch size,and using cosine annealing learning rate scheduler.

A. Classification

In the Tables V and in VI we showcase our results of our model on CT-scans using ResNet50 and InceptionResNetV2 respectively. The model is evaluated based on accuracy, F1 score, recall, precision, and specificity. Because both data sets, particularly the X-ray one, are skewed, relying just on accuracy might lead to misleading conclusions. We train each classifier 100 times using characteristics derived from pretrained CNNs in order to achieve statistically valid results. The performance measurements are calculated for each run. It should be emphasised that those figures were derived without the use of PCA for all classifiers that had been trained 100 times (all features passed to classifiers).

The Tables VII and VIII represents the scores of our model on X-ray Dataset. We train and assess each classifier 100 times to compare alternative feature extraction architectures in depth. Then we use the average of all predictions to get a credible sample label estimate. The accuracy, f1 score, precision, sensitivity, and specificity of the results are then determined. The reported values are expressed as a percentage. After comparing all models, we discovered that none of them outperformed the others in the majority of circumstances. For each model, linear SVM also produces the best results. We can also see that the ResNe50 outperforms the InceptionResNetV2 in both CT-scan and X-ray datasets. We also need to look at the Uncertainity quantification

Model	Classification	Accuracy	F1 score	Recall	Precision	Specificity
ResNet50	SVC (“Polynomial”)	91.23783763	91.23783763	91.45333333	91.89701293	92.09873621
	SVC (“RBF”)	92.34310992	92.34310992	92.38333333	93.12983043	93.46781230
	Random Forest	89.97354672	89.97354672	88.66666667	89.78637982	90.00962121
	Neural Network	95.76436721	95.76436721	95.5	96.67379291	96.97361231
	KNN	90.23123423	90.23123423	89.8333333	90.37599391	90.78112342
	Adaboost	87.63692712	87.63692712	85.1333333	86.59302893	86.91341743
	Naive Bayes	90.3333333	90.3333333	90.66666667	91.94809093	92.00341674
	XGBoost	94.0833333	94.0833333	93.5	94.98098012	95.16934313

TABLE V: Table showing the scores on Large CT-scan dataset using the ResNet50 model.

Model	Classification	Accuracy	F1 score	Recall	Precision	Specificity
InceptionResNetV2	SVC (“Polynomial”)	79.25	78.36192873	75.12987231	77.29790892	81.29872980
	SVC (“RBF”)	80.125	79.12878890	79.09281923	79.28978712	82.98701293
	Random Forest	81.5	80.87465823	79.99098211	80.12798790	83.82789072
	Neural Network	85.4375	84.83873847	83.44909812	84.00982131	87.18731291
	KNN	81.125	80.28773711	79.12870192	77.18912731	82.89379842
	Adaboost	81.4375	80.12734721	80.83910212	75.12837912	83.28972131
	Naive Bayes	45.5	53.34834789	71.28731789	41.76187236	50.23816231
	XGBoost	84.25	83.34893741	83.12809821	84.29732018	87.21379871

TABLE VI: Table showing the scores on Large CT-scan dataset using the InceptionResNetV2.

associated with the scores of each classification model. For example, take Adaboost the scores on both the datasets are exceptional but the uncertainty is near to 1 which is very bad. So, we have to consider both scores and uncertainty in order to rely on a particular model.

B. Segmentation

In this section we present our results of our two segmentation methodologies. The Table IX shows the results of transfer learning technique we used in segmentation model. The Table X shows the results of segmentation model using U-Net framework. In the Table IX we have used Efficient Net b[1-7] as our pretrained models. The X-ray and CT-images were given as input for the pretrained model, and extracted the features from the last layer by disconnecting fully connected layers in the Efficient Net. By using Efficient Net b7 we got Dice coefficient as 0.76 and IOU

Model	Classification	Accuracy	F1 score	Recall	Precision	Specificity
ResNet50	SVC (“Polynomial”)	85.83333333	85.12938791	85.12938791	85.12938791	83.67587667
	SVC (“RBF”)	86.83333333	85.82398213	85.82398213	85.82398213	82.93749013
	Random Forest	79.16666667	78.34309873	78.34309873	78.34309873	75.23792817
	Neural Network	88.5	88.10982838	88.10982838	88.10982838	87.29013821
	KNN	80.16666667	79.78190823	79.78190823	79.78190823	78.21831098
	Adaboost	79.5	78.23982730	78.23982730	78.23982730	79.21397830
	Naive Bayes	75.16666667	73.73897901	73.73897901	73.73897901	72.23981290
	XGBoost	87.83333333	85.34310809	85.34310809	85.34310809	85.09830129

TABLE VII: Table showing the scores on Large X-ray dataset using the ResNet50 model.

Model	Classification	Accuracy	F1 score	Recall	Precision	Specificity
InceptionResNetV2	SVC (“Polynomial”)	75.83333333	75.23098130	75.23098130	75.23098130	75.23098130
	SVC (“RBF”)	76.83333333	76.12938214	76.12938214	76.12938214	76.12938214
	Random Forest	78.16666667	77.31840123	77.31840123	77.31840123	77.31840123
	Neural Network	80.66666667	78.39048101	78.39048101	78.39048101	78.39048101
	KNN	73.33333333	71.09231908	71.09231908	71.09231908	71.09231908
	Adaboost	69.67777777	65.21398092	65.21398092	65.21398092	65.21398092
	Naive Bayes	62.33333333	59.21930934	59.21930934	59.21930934	59.21930934
	XGBoost	79.16666667	76.34789341	76.34789341	76.34789341	76.34789341

TABLE VIII: Table showing the scores on Large X-ray dataset using the InceptionResNetV2.

(Intersection over Union) as 0.612. In the Table X, we have used general U-Net framework along with various hyperparameters. Regarding hyperparameters we considered Block size, epochs, batch size, learning rate. By using block size as 6, epochs as 500, batch size as 64, learning rate as cosine annealing learning rate scheduler with initialization value of $1 * 10^{-5}$ we got our best results of Dice coefficient 0.86 and IOU (Intersection over Union) 0.754. On comparing both the results on 2 different datasets we can say that the U-Net framework is performing better than the transfer learning technique using EfficientNet in this scenario.

VIII. COMPARISON TO STATE-OF-THE-ART METHODS

We compare our model with the state-of-the-art methods on the COVID-19 Dataset. Tables XI ,XIIs hows the comparision results with U-Net [16], U-Net+[17], Mask-RCNN [18], BlitzNet [19],

Model	Type	Epochs	Dice Coeff	IOU
Transfer Learning	Efficient Net b1	400	0.72	0.5625
	Efficient Net b2	400	0.74	0.587302
	Efficient Net b3	400	0.69	0.526718
	Efficient Net b4	400	0.7	0.538462
	Efficient Net b5	400	0.75	0.6
	Efficient Net b6	400	0.73	0.574803
	Efficient Net b7	400	0.76	0.612903

TABLE IX: Table showing the results of Transfer Learning technique used in segmentation model.

Yolact [20], SOLO[21]. Codes of these methods are publicly available, and we follow the authors' instructions to retrain the models on the COVID-19 dataset.

Model	Size	Epochs	Learning Rate	Dice Coeff	IOU
U-Net	4	100	0.0005	0.75	0.6
	4	200	0.0005	0.76	0.612903
	4	300	0.0005	0.75	0.6
	4	400	0.0005	0.78	0.639344
	4	500	0.0005	0.8	0.666667
	6	100	0.0001	0.72	0.5625
	6	200	0.0001	0.73	0.574803
	6	300	0.0001	0.75	0.6
	6	400	0.0001	0.78	0.639344
	6	500	0.0001	0.81	0.680672
	6	100	0.0005	0.76	0.612903
	6	200	0.0005	0.77	0.626016
	6	300	0.0005	0.79	0.652893
	6	400	0.0005	0.81	0.680672
	6	500	0.0005	0.82	0.694915
	4	100	cosine(5* 10ˆ-5)	0.69	0.526718
	4	200	cosine(5* 10ˆ-5)	0.69	0.526718
	4	300	cosine(5* 10ˆ-5)	0.71	0.550388
	4	400	cosine(5* 10ˆ-5)	0.71	0.550388
	4	500	cosine(5* 10ˆ-5)	0.72	0.5625
	6	100	cosine(1* 10ˆ-5)	0.85	0.739130
	6	200	cosine(1* 10ˆ-5)	0.88	0.785714
	6	300	cosine(1* 10ˆ-5)	0.84	0.724137
	6	400	cosine(1* 10ˆ-5)	0.89	0.801801
	6	500	cosine(1* 10ˆ-5)	0.90	0.818181
	6	100	cosine(5* 10ˆ-5)	0.77	0.626016
	6	200	cosine(5* 10ˆ-5)	0.78	0.639344
	6	300	cosine(5* 10ˆ-5)	0.76	0.612903
	6	400	cosine(5* 10ˆ-5)	0.74	0.587302
	6	500	cosine(5* 10ˆ-5)	0.72	0.5625

TABLE X: This Table shows the results of segmentation model using U-net framework.

Method	IOU	Dice	Pixel Accuracy
U-Net	66.8	78.9	77.7
Attention U-Net	69.7	82.1	81.4
U-Net+	67.8	80.2	78.4
Mask-RCNN	67.2	80.7	78.5
BlitzNet	69.5	81.1	81.7
Yolact	58.4	74.2	69.9
SOLO	70.1	83.0	82.4
SRGNet	71.1	83.0	83.2
Mod U-Net Attention (Ours)	81.1	90.4	88.6

TABLE XI: Segmentation and detection comparison among our Modified U-Net Attention model and other state-of-the-art methods on the COVID-19 dataset - I. [4]

Method	IOU	Dice	Pixel Accuracy
U-Net	52.2	68.6	69.8
Attention U-Net	59.1	74.3	72.5
U-Net+	53.2	69.5	65.8
Mask-RCNN	53.9	70.1	71.2
BlitzNet	65.8	79.4	77.3
Yolact	48.3	65.2	66.8
SOLO	61.8	76.4	75.3
SRGNet	66.8	80.1	78.5
Mod U-Net Attention (Ours)	75.7	86.2	83.4

TABLE XII: Segmentation and detection comparison among our Modified U-Net Attention model and other state-of-the-art methods on the COVID-19 dataset - II. [5]

IX. COMPARISON TO OTHER PAPERS

We used a common dataset [31] to simulate our model. The outcomes of our model were then compared to those of other previously published works. Table XIII demonstrates that our model performs better than any other model we have looked at.

Model	Dice	IOU
Paper-I [32]	77.1	66.86
Paper-II [33]	80.4	67.2
Paper-III [34]	75.8	61.3
Paper-IV [35]	83.1	71.08
Paper-V [?]	70.3	54.2
Ours	90.4	81.1

TABLE XIII: In this table we present the comparision results

X. CONCLUSION

We have proposed an effective model to jointly detect and segment the lesions of COVID-19. For segmentation, our model simply takes into consideration the covid positive pictures. We can save a lot of processing power and time this way. We don't have to worry about the dataset being imbalanced because we factored in weighted boundary loss and binary cross entropy dice loss. The experimental results have demonstrated the superior performance of our model. However, there are still some limitations worth noting.

To begin with, the identification and segmentation of the COVID-19 lesion region are significantly reliant on clinical expert annotation, which is time-consuming and influenced by existing information. Second, COVID-19's afflicted region has imaging characteristics with other pneumonia viruses. We don't include other viral pneumonia for comparison because each cause hasn't been confirmed. Finally, CT scans obtained by various devices will obstruct the algorithm's advancement.

In the future, we want to enhance the number of COVID-19 annotated cases and invite more experts to verify the accuracy of the annotated data. We also intend to incorporate data on pneumonia caused by additional viruses and retrain the data to increase the algorithm's resilience.

XI. REFERENCES

[1] G. Yadav, S. Maheshwari and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 2392-2397, doi: 10.1109/ICACCI.2014.6968381.

[2] https://www.kaggle.com/datasets/maedemaftouni/large-covid19-ct-slice-dataset

[3] https://www.kaggle.com/datasets/ajeyakrishna/covid-xray-dataset?select=COVID

[4] https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans

[5] https://www.kaggle.com/datasets/maedemaftouni/large-covid19-ct-slice-dataset

[6] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition" in arXiv:1512.03385, 2015.

[7] C. Szegedy et al., "Going deeper with convolutions", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1-9, Jun. 2015.

[8] A. Shamsi et al., "An Uncertainty-Aware Transfer Learning-Based Framework for COVID-19 Diagnosis," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 4, pp. 1408-1417, April 2021, doi: 10.1109/TNNLS.2021.3054306.

[9] Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., Dong, Y. (2021). The ensemble deep learning model for novel COVID-19 on CT images. Applied soft computing, 98, 106885.https://doi.org/10.1016/j.asoc.2020.106885

[10] Tenório, G.L., Villalobos, C.M., Mendoza, L.F., Silva, E.C., Caarls, W. (2019). Improving Transfer Learning Performance: An Application in the Classification of Remote Sensing Data. ICAART.

[11] Wu, Y., Gao, S., Mei, J., Xu, J., Fan, D., Zhao, C., Cheng, M. (2021). JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. IEEE Transactions on Image Processing, 30, 3113-3126.

[12] J. van Amersfoort, L. Smith, Y. Whye Teh and Y. Gal, "Simple and scalable epistemic uncertainty estimation using a single deep deterministic neural network" in arXiv:2003.02037, Mar. 2020

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition" in arXiv:1409.1556, 2014.

[14] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition" in arXiv:1512.03385, 2015.

[15] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan China", Lancet, vol. 395, pp. 497-506, May 2020.

[16] Ronneberger O., Fischer P., Brox T. International Conference on Medical Image Computing and Computer-assisted Intervention; 2015. U-Net: convolutional networks for biomedical image segmentation; pp. 234–241.

[17] . Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer; 2018. UNet++: a nested U-Net architecture for medical image segmentation; pp. 3–11.

[18] He K., Gkioxari G., Dollár P., Girshick R. Proceedings of the IEEE International Conference on Computer Vision. 2017. Mask R-CNN; pp. 2961–2969.

[19] Dvornik N., Shmelkov K., Mairal J., Schmid C. Proceedings of the IEEE International Conference on Computer Vision. 2017. BlitzNet: a real-time deep network for scene understanding; pp. 4154–4162.

[20] Bolya D., Zhou C., Xiao F., Lee Y.J. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. YOLACT: real-time instance segmentation; pp. 9157–9166.

[21] Wang X., Kong T., Shen C., Jiang Y., Li L. European Conference on Computer Vision. Springer; 2020. SOLO: segmenting objects by locations; pp. 649–665.

[22] Wu, J., Zhang, S., Li, X., Chen, J., Xu, H., Zheng, J., Gao, Y., Tian, Y., Liang, Y., Ji, R. (2021). Joint segmentation and detection of COVID-19 via a sequential region generation network. Pattern recognition, 118, 108006. https://doi.org/10.1016/j.patcog.2021.108006

[23] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, Ismail Ben Ayed, Boundary loss for highly unbalanced segmentation.

[24] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention U-Net: learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).

[25] S. Ren, K. He, R. Girshick, J. Sun Faster R-CNN: towards real-time object detection with region proposal networks IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2016), pp. 1137-1149

[26] Z. Cai, N. Vasconcelos Cascade R-CNN: delving into high quality object detection Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018), pp. 6154-6162

[27] K. He, R. Girshick, P. Dollár Rethinking imagenet pre-training Proceedings of the IEEE International Conference on Computer Vision (2019), pp. 4918-4927

[28] N. Khehrah, M.S. Farid, S. Bilal, M.H. Khan Lung nodule detection in CT images using statistical and shape-based features J. Imaging, 6 (2) (2020), p. 6

[29] H. Xie, D. Yang, N. Sun, Z. Chen, Y. Zhang Automated pulmonary nodule detection in CT images using deep convolutional neural networks

[30] L. Wang, A. Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images, arXiv preprint arXiv:2003.09871 (2020).

[31] (2020) Covid-19 ct segmentation dataset. URL http:// medicalsegmentation.com/covid19/

[32] João O. B. Diniz et.al,Segmentation and quantification of COVID-19 infections in CT using pulmonary vessels extraction and deep learning

[33] Müller D, Soto-Rey I, Kramer F. Robust chest CT image segmentation of COVID-19 lung infection based on limited data. Inform Med Unlocked. 2021;25:100681. doi: 10.1016/j.imu.2021.100681. Epub 2021 Jul 27. PMID: 34337140; PMCID: PMC8313817.

[34] Jianxiong Zhang et.al, Semantic segmentation of COVID-19 lesions with a multiscale dilated convolutional network

[35] Zhou T, Canu S, Ruan S. Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. Int J Imaging Syst Technol. 2021 Mar;31(1):16-27. doi: 10.1002/ima.22527. Epub 2020 Nov 24. PMID: 33362345; PMCID: PMC7753491.